

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Dalam konteks pendidikan tinggi di Indonesia, karya ilmiah merupakan persyaratan yang harus dipenuhi oleh mahasiswa untuk mendapatkan gelar Sarjana Strata-1. Setiap perguruan tinggi memiliki ketentuan tersendiri dalam penyelesaian tugas akhir, yang dapat berupa skripsi, tesis, jurnal ilmiah, artikel, prototipe, dan berbagai bentuk lainnya. Salah satu komponen penting dalam karya ilmiah adalah abstrak, yang berfungsi sebagai ringkasan singkat dari keseluruhan isi penelitian. Abstrak ini menyajikan pokok-pokok penting dari penelitian yang kemudian akan dijelaskan secara lebih lengkap dalam isi jurnal ilmiah (Indriati, 2023).

Berdasarkan Jurnal Ilmiah Teknologi Informasi dan Robotika yang selalu bertambah setiap tahunnya, maka seharusnya terdapat informasi dari kumpulan dokumen jurnal tersebut. Meskipun jumlah dokumen jurnal terus meningkat setiap tahun, pada umumnya, belum ada penelitian lanjutan yang dapat memberikan ringkasan informasi yang memadai dari dokumen-dokumen tersebut. Oleh karena itu, dengan penambahan jurnal setiap tahun dan pertumbuhan jumlah dokumen jurnal, penting untuk mengadopsi metode *text mining* untuk mengeksplorasi dan mengelola informasi yang terdapat dalam kumpulan dokumen jurnal tersebut. (Fauzi, 2022).

Saat ini, banyak perusahaan mengadopsi sistem data *mining* untuk melakukan segmentasi pasar. Data *mining* merupakan proses eksplorasi dan penemuan pengetahuan dari kumpulan data yang tersedia. Proses ini melibatkan

penggunaan berbagai metode seperti analisis statistik, penerapan konsep matematika, kecerdasan buatan, dan teknik pembelajaran mesin untuk mengambil informasi dan pengetahuan yang bermanfaat dari berbagai kumpulan data besar. Hal ini memungkinkan perusahaan untuk mengoptimalkan strategi pemasaran dan mengidentifikasi peluang bisnis yang berpotensi (Sano, 2019; Turban, Aronson, & Liang, 2019).

Data *mining* tidak hanya digunakan dalam analisis perusahaan, melainkan juga dalam berbagai bidang lain. Contoh aplikasinya meliputi telekomunikasi untuk menganalisis transaksi, keuangan untuk mendeteksi transaksi mencurigakan, dan penjelajahan internet untuk menganalisis perilaku pelanggan. Selain itu, terdapat bidang serupa seperti penambangan teks, yang fokus pada penemuan pola dalam data tekstual besar. Tujuannya adalah untuk menemukan informasi yang berguna untuk berbagai tujuan (Putra, 2019; Diaz, 2019; Ronen dan Sanger, 2020).

Salah satu peran dari penambangan data dan penambangan teks adalah pengelompokan (*clustering*). Pengelompokan merupakan metode dalam penambangan data atau teks yang tidak memerlukan supervisi. Metode ini tidak membutuhkan data latihan atau label target keluaran. Clustering terbagi menjadi dua kategori utama: clustering hierarki dan clustering non-hierarki. Pendekatan hierarki melibatkan pengelompokan dua atau lebih objek yang memiliki kesamaan yang paling tinggi terlebih dahulu. Langkah ini kemudian diulangi dengan menambahkan objek tambahan ke dalam kelompok tersebut. Ini menghasilkan struktur pohon hierarki di mana hubungan antara objek diwakili sebagai tingkat (level) pada pohon. Di sisi lain, pengelompokan non-hierarkis dimulai dengan

menentukan jumlah *cluster* yang diinginkan terlebih dahulu, dan kemudian objek dikelompokkan berdasarkan kriteria tertentu tanpa memperhatikan struktur hierarki (Agusta, 2019).

Dalam Jurnal Ilmiah Teknologi Informasi dan Robotika, permasalahan yang muncul adalah belum adanya pengetahuan tentang jumlah klaster topik penelitian yang ada. Penerapan teknik *clustering* pada topik penelitian diperlukan untuk memahami tren topik yang sedang berkembang. Pengelompokan topik bisa dilakukan secara manual oleh siswa, tetapi membutuhkan waktu yang cukup lama. Untuk mengatasi kendala tersebut, penggunaan komputer dengan teknik pemodelan dapat menjadi solusi yang lebih efisien. Secara umum, terdapat dua jenis penulisan esai, salah satunya terfokus pada konteks teknologi informasi dan robotika.

Pemodelan topik merupakan metode pengelompokan non-hierarki yang secara otomatis mengorganisir topik yang muncul dari data untuk membentuk kelompok topik terkait. Pendekatan ini bertujuan untuk mengatasi tantangan yang sering dihadapi saat menganalisis jurnal ilmiah di bidang teknologi informasi dan robotika. Terdapat berbagai metode pemodelan topik, termasuk Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA), Latent Dirichlet Allocation (LDA), dan Hierarchical Dirichlet Process (HDP). LSA merupakan teknik pemrosesan bahasa alami yang bertujuan untuk mengungkap makna dengan menciptakan representasi vektor teks. PLSA, yang diperkenalkan oleh Puzicha dan Hofmann, adalah pendekatan untuk mengindeks dokumen secara otomatis menggunakan model kelas statistik laten, dengan upaya untuk meningkatkan aspek probabilistik dari metode LSA. LDA, yang dikembangkan oleh Blei, Ng, dan Theo,

adalah peningkatan dari PLSA yang menggabungkan elemen pemodelan campuran untuk menangkap interaksi antara kata dan dokumen. HDP, di sisi lain, adalah model untuk data yang dapat dikelompokkan di mana setiap item data dapat termasuk ke dalam beberapa kelompok yang berbeda, seperti yang terjadi dalam model topik di mana kata-kata dikelompokkan ke dalam dokumen yang mungkin memiliki topik-topik yang saling tumpang tindih.

Dalam penelitian ini, memanfaatkan model Bayesian nonparametrik yang dikenal sebagai Hierarchical Dirichlet Process (HDP) untuk mengelompokkan masalah dengan beberapa kumpulan data. HDP memungkinkan pemodelan setiap kelompok data sebagai gabungan komponen terbuka yang secara otomatis dihasilkan dari model tersebut. HDP memiliki kemampuan untuk memisahkan komponen di antara kelompok-kelompok tersebut, memungkinkan penggunaan pemodelan ketergantungan antar kelompok secara efisien. Hasil pengujian pada tiga set dokumen menunjukkan bahwa HDP menunjukkan kinerja yang lebih unggul dan efisien dibandingkan dengan model sebelumnya. Sehingga, HDP digunakan dalam Jurnal Ilmiah Teknologi Informasi dan Robotika (JIFTI) sebagai bagian dari penelitian ini.

## **1.2 Rumusan Masalah**

1. Bagaimana proses implementasi *topic modelling* menggunakan metode *Hierarchical Dirichlet Process* (HDP) pada data abstrack Jurnal Ilmiah Teknologi Informasi dan Robotika (JIFTI)?

2. Apa saja topik yang muncul dari pemodelan topik pada abstrak Jurnal Ilmiah Teknologi Informasi dan Robotika menggunakan metode Hierarchical Dirichlet Process (HDP)?

### **1.3 Batasan Penelitian**

Berikut adalah beberapa batasan masalah dalam penelitian ini:

1. Sumber data yang digunakan berasal dari *website* Jurnal Ilmiah Teknologi Informasi dan Robotika.
2. Studi kasus yang diteliti adalah abstrak penelitian pada Program Studi Teknik Informatika pada Jurnal Ilmiah Teknologi Informasi dan Robotika (JIFTI) tahun 2019-2022, dengan total 42 abstrak.
3. Metode yang diterapkan dalam *topic modelling* adalah *Hierarchical Dirichlet Process* (HDP).
4. Pengambilan data abstrak dilakukan secara manual dengan menggunakan *tools* Chrome.

### **1.4 Tujuan Penelitian**

1. Untuk mengetahui proses implementasi *topic modelling* menggunakan metode *Hierarchical Dirichlet Process* (HDP) pada data abstrak Jurnal Ilmiah Teknologi Informasi dan Robotika (JIFTI)
2. Melakukan identifikasi dan mengetahui topik-topik yang muncul dari pemodelan topik pada abstrak Jurnal Ilmiah Teknologi Informasi dan Robotika dengan menggunakan metode *Hierarchical Dirichlet Process* (HDP).

## 1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah:

1. Manfaat teoritis:
  - a. Informasi tambahan mengenai penerapan metode *Hierarchical Dirichlet Process* (HDP) dalam proses klasifikasi topik penelitian di Jurnal Ilmiah Teknologi Informasi dan Robotika.
  - b. Memperoleh pemahaman yang lebih mendalam tentang penambangan data, termasuk teknik-teknik seperti penambangan teks dan metode *Hierarchical Dirichlet Process* (HDP), yang secara khusus digunakan untuk analisis topik dalam teks.
2. Manfaat praktis:
  - a. Memberikan bantuan kepada Jurnal Ilmiah Teknologi Informasi dan Robotika dalam mengidentifikasi berbagai topik penelitian yang relevan.
  - b. Pemodelan yang terdapat dalam penelitian ini memiliki potensi untuk mengurangi waktu, upaya, dan biaya yang diperlukan.